



# CAR culadora

**Descubra quanto vale seu  
seminovo ou usado** de uma forma  
mais completa que a Tabela Fipe

**27 milhões**

**59 milhões**



**1 CARRO PARA CADA 4 BRASILEIROS**

**22% do PIB da indústria e por  
4% do PIB total, empregando  
mais de 1 milhão de pessoas**

# FOLHA DE S.PAULO



## **Produção de veículos acumula queda de 13,6% até abril**

---

10/05/2022

# SETOR AUTOMOTIVO

Fabricação e venda de peças de reposição

Serviços de manutenção

Seguros

Blindagem

Locação de veículos

Revenda de carros seminovos e usados

**14 milhões de veículos vendidos,  
crescimento de 2,2% em relação  
ao ano anterior e um faturamento  
de 380 Bilhões**

# Tabela Fipe

Marca

Modelo

Ano

# CARculadora

Marca

Modelo

Ano

Quilometragem

Itens Adicionais

Único Dono

Unidade Federativa

...

# CONSTRUÇÃO DO **MODELO**



# Aquisição de dados

Técnica de web scraping

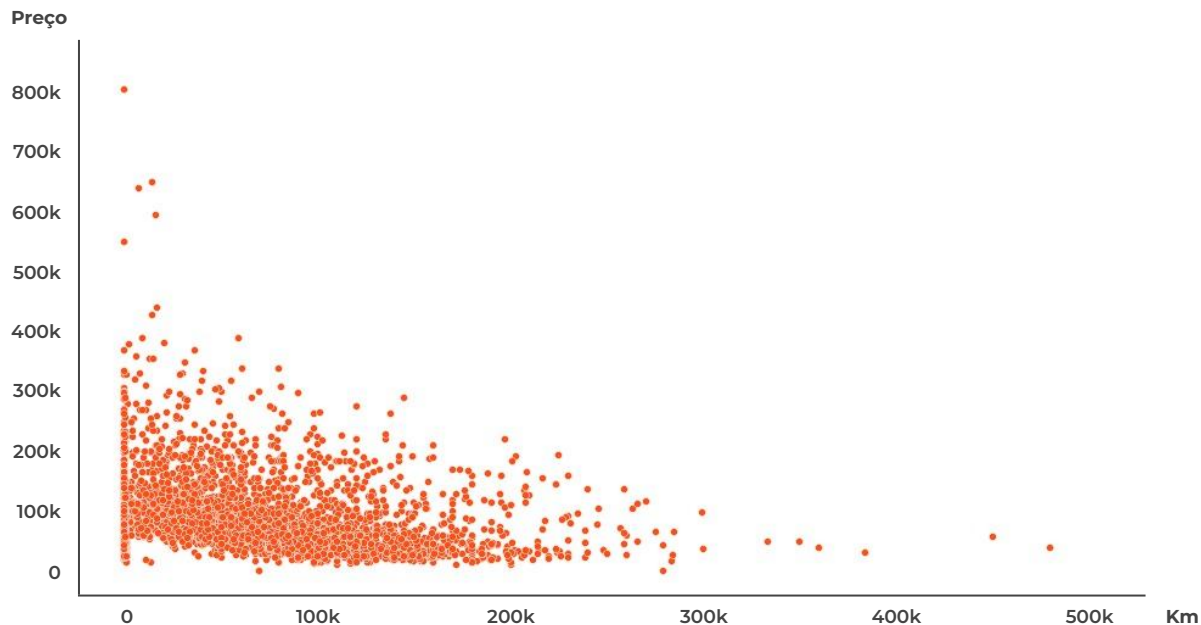
Mais de 5000 dados coletados



*“É a marca mais lembrada na hora de comprar ou vender carros online, são mais de 5 carros vendidos por minuto”*

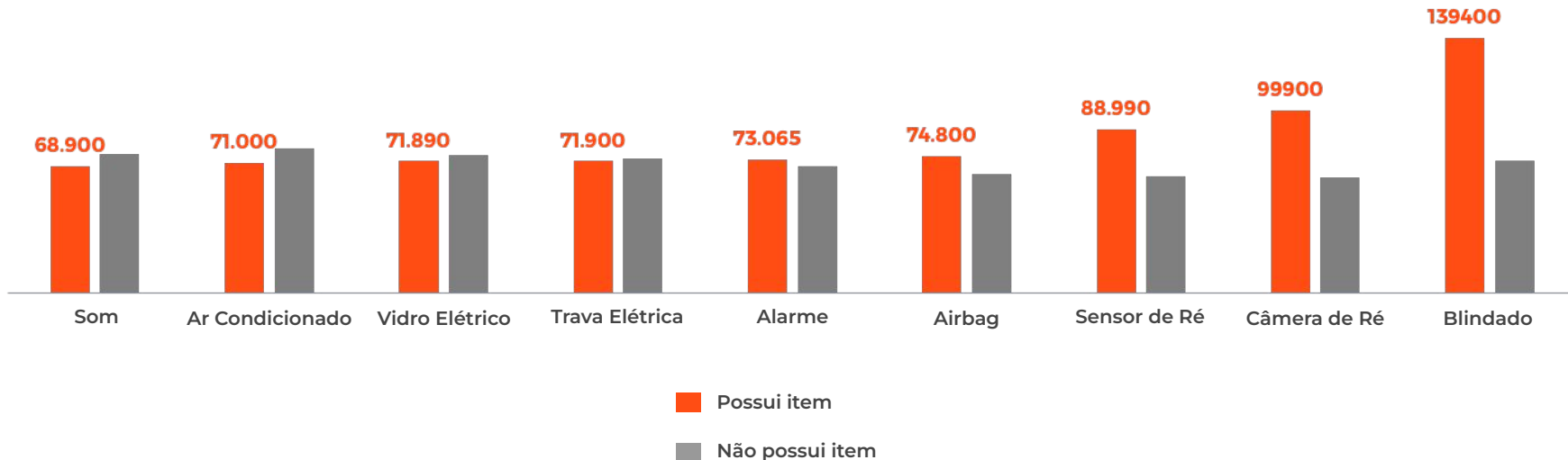
# Análise Exploratória de Dados (EDA)

## Preço x Quilometragem



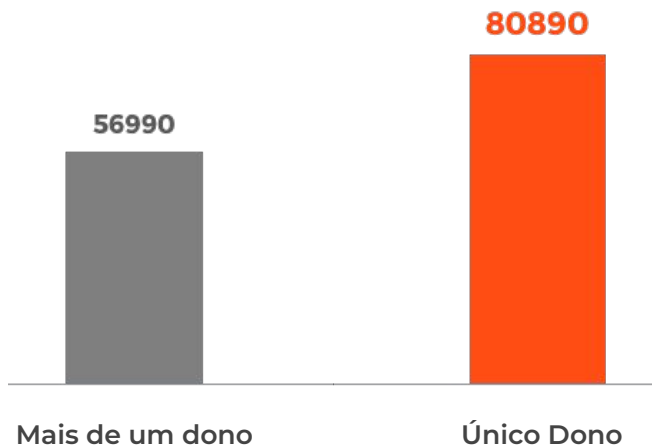
# Análise Exploratória de Dados (EDA)

## Preço x Itens Opcionais



# Análise Exploratória de Dados (EDA)

Preço x Único Dono



# Feature engineering

**KmPorDias:** quilômetros diários percorridos pelo veículo;

**NumOpcionais:** número total de itens opcionais;

**KmSqrt:** raiz quadrada da quilometragem do carro;

**PotenciaNum:** versão numérica da potência do automóvel.

# PROBLEMAS NA MODELAGEM

OBTER DADOS DA  
TABELA FIPE

DESCRIÇÃO DOS  
ANÚNCIOS OLX

RESULTADO MODELO  
VS BASELINE

# PROBLEMAS NA MODELAGEM

OBTER DADOS DA  
TABELA FIPE

DESCRIÇÃO DOS  
ANÚNCIOS OLX

RESULTADO MODELO  
VS BASELINE

Acessar API pública  
e gratuita para  
conseguir dados

# PROBLEMAS NA MODELAGEM

OBTER DADOS DA  
TABELA FIPE

DESCRIÇÃO DOS  
ANÚNCIOS OLX

RESULTADO MODELO  
VS BASELINE

**Eliminar anúncios  
destoantes, cerca de  
30% da base**



# PROBLEMAS NA MODELAGEM

OBTER DADOS DA  
TABELA FIPE

DESCRIÇÃO DOS  
ANÚNCIOS OLX

RESULTADO MODELO  
VS BASELINE

| Modelo            | RMSE     | MAPE   | RMSLE  |
|-------------------|----------|--------|--------|
| Tabela Fipe       | 11388,50 | 0,0813 | 0,1452 |
| Árvore de decisão | 32264,06 | 0,1811 | 0,3385 |
| Random Forest     | 24738,70 | 0,1404 | 0,2026 |
| XGBoost           | 25694,72 | 0,1368 | 0,2012 |

# Modelos Testados

**Árvore de decisão:** modelo que divide a base de dados de acordo com os valores de certas features, gerando **regras de decisão**

**Random Forest:** ensemble do tipo *bagging* baseado em árvores de decisão, baseado no “**conhecimento da maioria**”

**XGBoost:** ensemble do tipo *boosting* que utiliza árvores de decisão de forma serial, **minimizando o erro** do modelo anterior

# Métricas Utilizadas

**RMSE:** métrica muito popular que penaliza os erros maiores.  
Resultado na unidade da variável dependente (R\$)

**MAPE:** calcula o erro em termos de uma porcentagem em relação ao valor real. Não depende da escala dos valores

**RMSLE:** similar ao RMSE, porém mais robusta a outliers

Se não pode  
vencê-los,  
**junte-se a eles**

# Resultados utilizando a Fipe

| Árvore de<br>decisão | RMSE | MAPE | RMSLE |
|----------------------|------|------|-------|
| Treino               | 0    | 0    | 0     |

**PROBLEMA: OVERFITTING**

# O QUE FIZEMOS

OBTER MAIS DADOS

SELECIONAR FEATURES  
DE ENTRADA

SINTONIZAR  
HIPERPARÂMETROS

# O QUE FIZEMOS

OBTER MAIS DADOS

SELECIONAR FEATURES  
DE ENTRADA

SINTONIZAR  
HIPERPARÂMETROS

~~Demanda muito tempo e esforço~~

# O QUE FIZEMOS

OBTER MAIS DADOS

SELECIONAR FEATURES  
DE ENTRADA

SINTONIZAR  
HIPERPARÂMETROS

Exclusão de features com **pouca importância ou redundantes**

Criação de novas features baseadas nas **fraquezas da tabela Fipe**



# O QUE FIZEMOS

OBTER MAIS DADOS

SELECIONAR FEATURES  
DE ENTRADA

SINTONIZAR  
HIPERPARÂMETROS

**n\_estimators:** número de árvores de decisão utilizadas

**max\_depth:** profundidade máxima da árvore

**min\_samples\_split:** número mínimo de amostras para gerar uma nova divisão no nó

**max\_features:** proporção de features consideradas ao gerar uma divisão

**Random Forest Quantile: prever os preços com intervalos**

# Resultados utilizando a Fipe

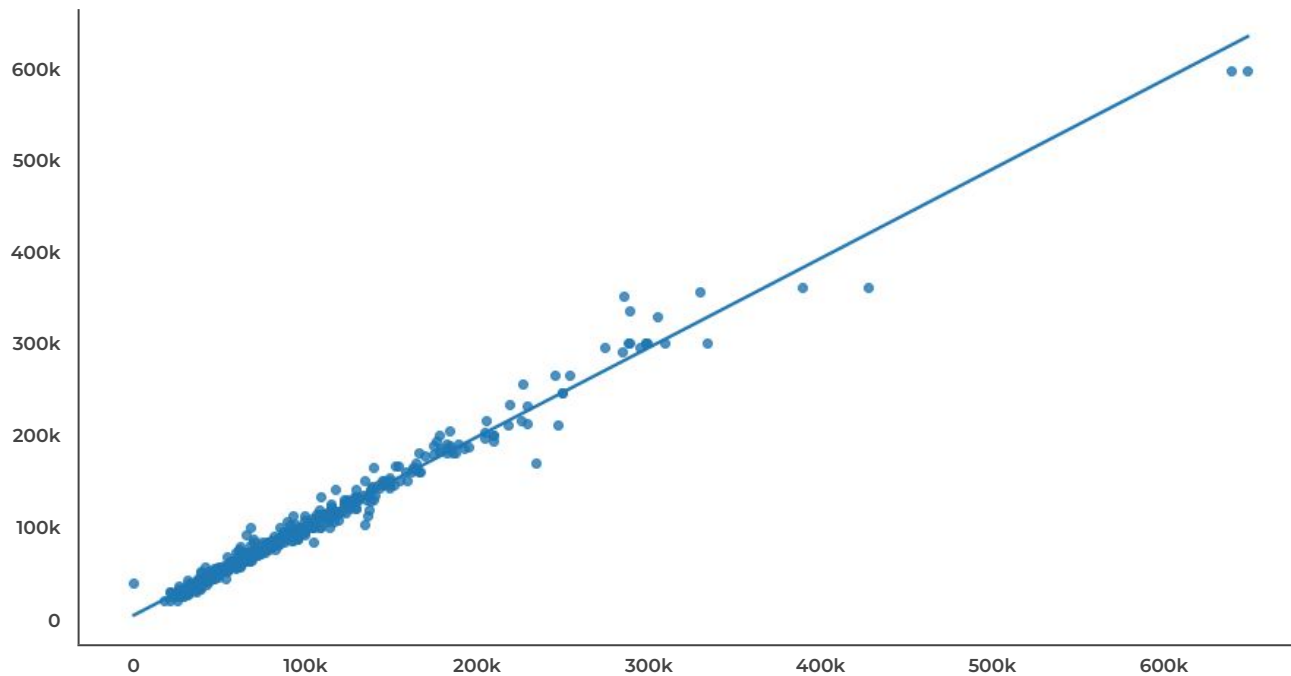
## Validação:

| Modelo      | RMSE     | MAPE   | RMSLE  |
|-------------|----------|--------|--------|
| Tabela Fipe | 11388,50 | 0,0813 | 0,1452 |
| RF Quantile | 11690,04 | 0,0783 | 0,1452 |

## Base de teste:

| Modelo      | RMSE     | MAPE   | RMSLE  |
|-------------|----------|--------|--------|
| Tabela Fipe | 10090,56 | 0,1512 | 0,1887 |
| RF Quantile | 9303,82  | 0,1459 | 0,1855 |

# Preço real versus predito



# R\$ 341

de erro a menos que a tabela Fipe,  
por automóvel, na base de teste

# R\$ 187k

de erro a menos que a tabela Fipe,  
em 549 veículos da base de teste

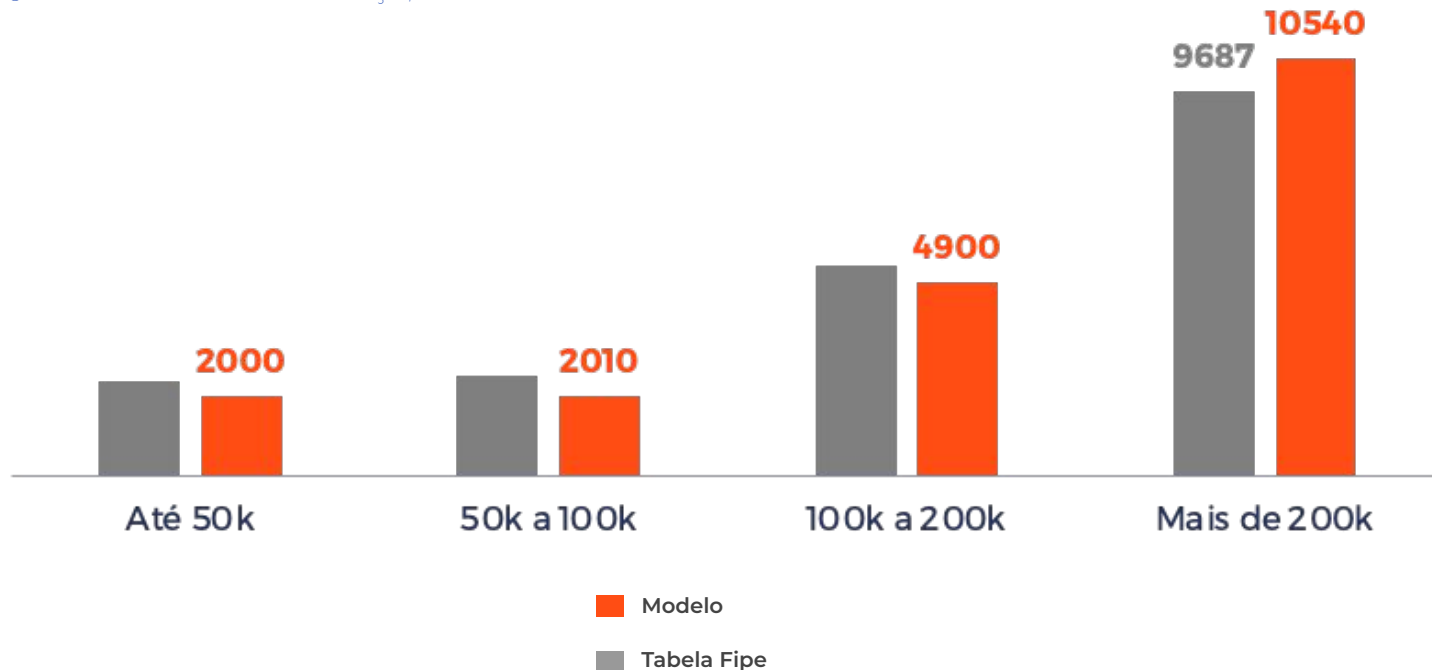
# 91%

dos valores da base de teste dentro dos  
intervalos determinados pelo modelo

# Onde ganhamos, onde nem tanto

## Diferença para o valor real (anúncio)

Quanto menor a diferença, melhor



# Comitê de Ética

Evitar distorções exacerbadas de preço por **unidade federal**

Tornar pública as **forças e fraquezas** do modelo

Aumentar a **transparência** da consulta em negociações

# DEPLOY

A thick, solid orange horizontal bar positioned directly beneath the word "DEPLOY".

🚗 Início

📄 CARculadora

📝 Análise



Tenha acesso aos preços reais de mercado! Entre com os dados de seu veículo abaixo

Marca

ACURA

Modelo

Integra GS 1.8

Ano/Combustível

1992 Gasolina

Flex



A tabela Fipe termina aqui. Mas a gente pode te oferecer uma previsão ainda melhor com as informações abaixo:

Tipo de carroceria

☒ Passeio ☐ Hatch ☐ Sedã ☐ SUV ☐ Pick-up ☐ Van/Utilitário ☐ Conversível

Estado

RJ

Final de placa

0

Quilometragem

Direção



Obrigado